CS 287 Lecture 12 (Fall 2019)
Kalman Filtering

Lecturer: Ignasi Clavera
Slides by Pieter Abbeel
UC Berkeley EECS

Many slides adapted from Thrun, Burgard and Fox, Probabilistic Robotics

Outline

- Gaussians
- Kalman filtering
- Extend Kalman Filter (EKF)
- Unscented Kalman Filter (UKF) [aka “sigma-point filter”]

Outline

- ***Gaussians***
- Kalman filtering
- Extend Kalman Filter (EKF)
- Unscented Kalman Filter (UKF) [aka “sigma-point filter”]

Multivariate Gaussians

$$p(x;\mu,\Sigma) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right)$$

$$\int \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right) dx = 1$$

For a matrix $A \in \mathbb{R}^{n\times n}$, $|A|$ denotes the determinant of A.

For a matrix $A \in \mathbb{R}^{n\times n}$, A^{-1} denotes the inverse of A, which satisfies $A^{-1}A = I = AA^{-1}$ with $I \in \mathbb{R}^{n\times n}$ the identity matrix with all diagonal entries equal to one, and all off-diagonal entries equal to zero.

Hint: often when trying to understand matrix equations, it's easier to first consider the special case of the dimensions of the matrices being one-by-one. Once parsing them that way makes sense, a good second step can be to parse them assuming all matrices are diagonal matrices. Once parsing them that way makes sense, usually it is only a small step to understand the general case.

Multivariate Gaussians

$$\mathsf{E}_X[X_i] = \int x_i p(x;\mu,\Sigma)dx = \mu_i$$

$$\mathsf{E}_X[X] = \int x p(x;\mu,\Sigma)dx = \mu$$

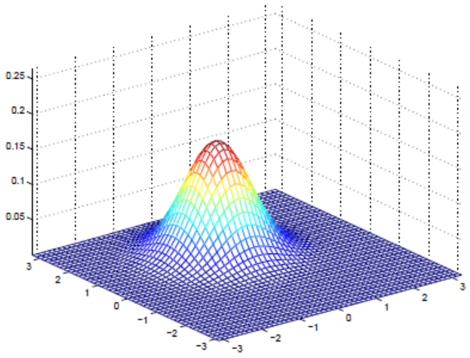
(integral of vector = vector of integrals of each entry)

$$\mathsf{E}_X[(X_i-\mu_i)(X_j-\mu_j)] = \int (x_i-\mu_i)(x_j-\mu_j)p(x;\mu,\Sigma)dx = \Sigma_{ij}$$

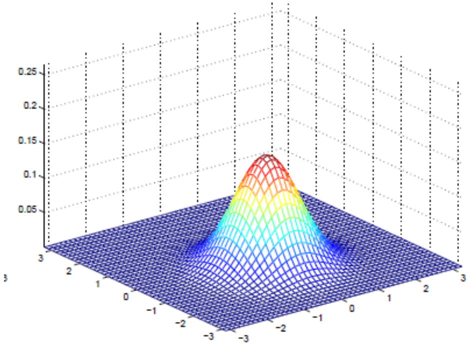
$$\mathsf{E}_X[(X-\mu)(X-\mu)^\top] = \int [(X-\mu)(X-\mu)^\top p(x;\mu,\Sigma)dx = \Sigma$$

(integral of matrix = matrix of integrals of each entry)

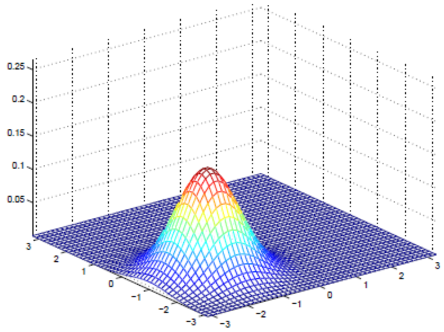
Multivariate Gaussians: Examples



- μ = [1; 0]
- Σ = [1 0; 0 1]

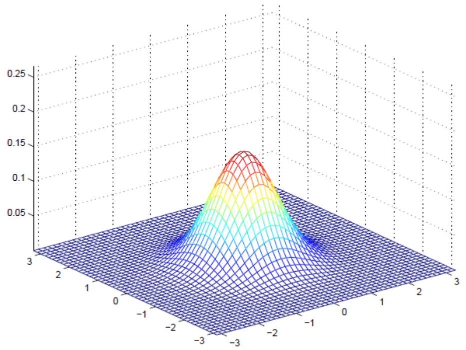


- μ = [-.5; 0]
- Σ = [1 0; 0 1]

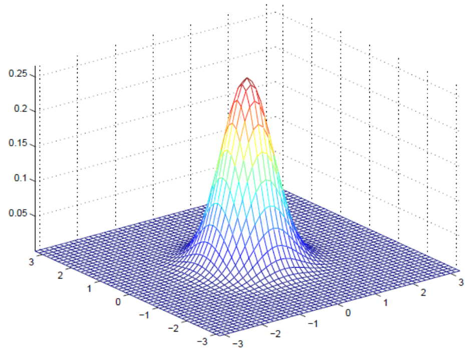


- μ = [-1; -1.5]
- Σ = [1 0; 0 1]

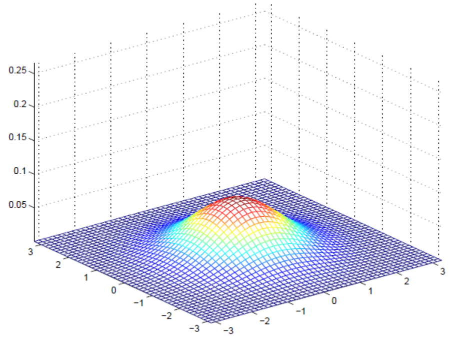
Multivariate Gaussians: Examples



- μ = [0; 0]
- Σ = [1 0 ; 0 1]

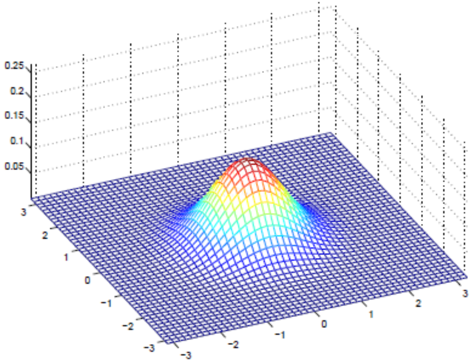


- μ = [0; 0]
- Σ = [.6 0 ; 0 .6]

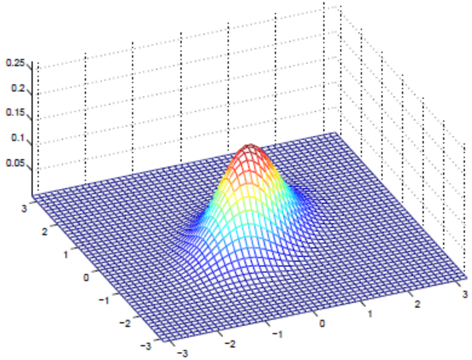


- μ = [0; 0]
- Σ = [2 0 ; 0 2]

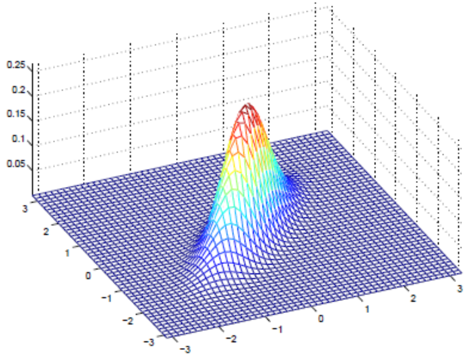
Multivariate Gaussians: Examples



- μ = [0; 0]
- Σ = [1 0; 0 1]

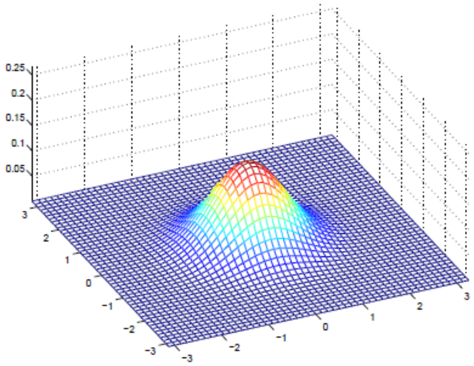


- μ = [0; 0]
- Σ = [1 0.5; 0.5 1]

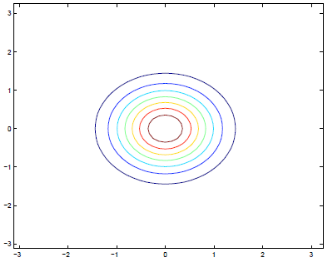


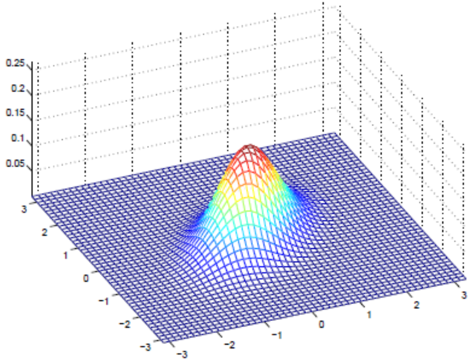
- μ = [0; 0]
- Σ = [1 0.8; 0.8 1]

Multivariate Gaussians: Examples

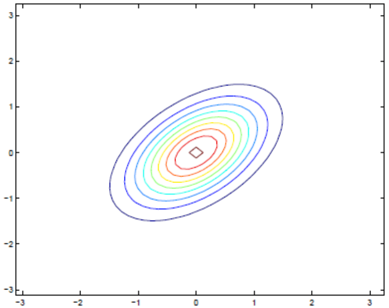


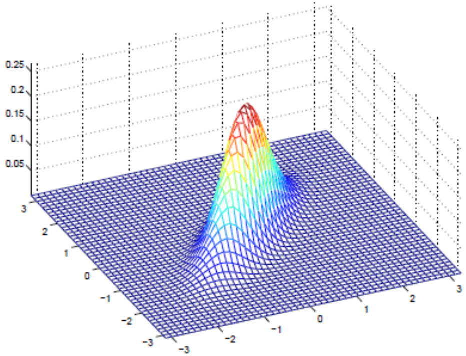
- $\mu = [0; 0]$
- $\Sigma = [1 \;\; 0; 0 \;\; 1]$



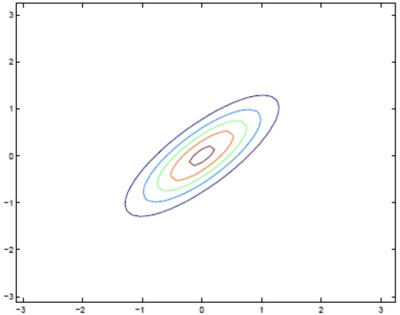


- $\mu = [0; 0]$
- $\Sigma = [1 \;\; 0.5; 0.5 \;\; 1]$

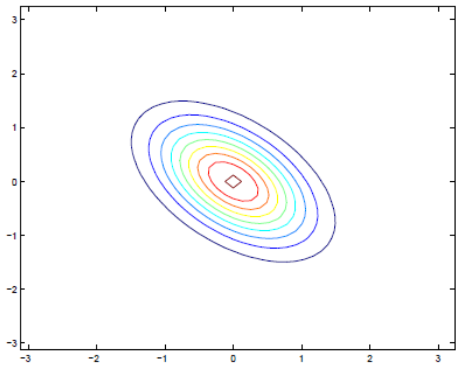




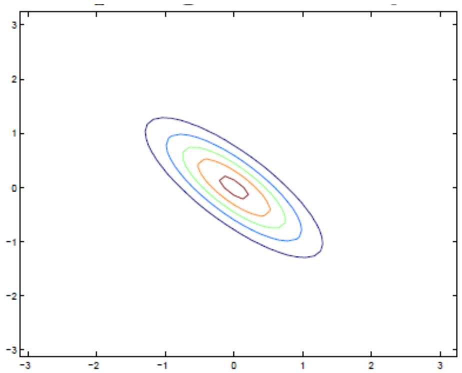
- $\mu = [0; 0]$
- $\Sigma = [1 \;\; 0.8; 0.8 \;\; 1]$



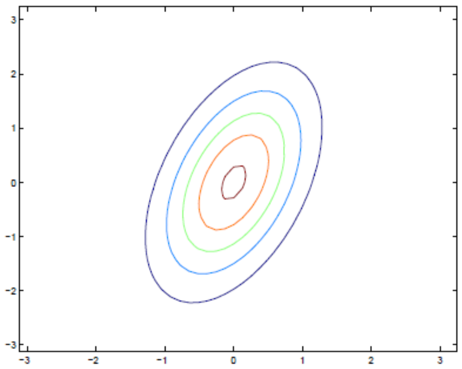
Multivariate Gaussians: Examples



- μ = [0; 0]
- Σ = [1 -0.5 ; -0.5 1]



- μ = [0; 0]
- Σ = [1 -0.8 ; -0.8 1]



- μ = [0; 0]
- Σ = [3 0.8 ; 0.8 1]

Partitioned Multivariate Gaussian

- Consider a multi-variate Gaussian and partition random vector into (X, Y).

$$\mathcal{N}(\mu, \Sigma) = \mathcal{N}\left(\begin{bmatrix}\mu_X \\ \mu_Y\end{bmatrix}, \begin{bmatrix}\Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY}\end{bmatrix}\right)$$

$$p(\begin{bmatrix}x \\ y\end{bmatrix}; \mu, \Sigma) = \frac{1}{(2\pi)^{(n/2)}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\left(\begin{bmatrix}x \\ y\end{bmatrix} - \begin{bmatrix}\mu_X \\ \mu_Y\end{bmatrix}\right)^\top \begin{bmatrix}\Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY}\end{bmatrix}^{-1} \left(\begin{bmatrix}x \\ y\end{bmatrix} - \begin{bmatrix}\mu_X \\ \mu_Y\end{bmatrix}\right)\right)$$

$$\begin{aligned}
\mu_X &= \mathrm{E}_{(X,Y)\sim\mathcal{N}(\mu,\Sigma)}[X] \\
\mu_Y &= \mathrm{E}_{(X,Y)\sim\mathcal{N}(\mu,\Sigma)}[Y] \\
\Sigma_{XX} &= \mathrm{E}_{(X,Y)\sim\mathcal{N}(\mu,\Sigma)}[(X-\mu_X)(X-\mu_X)^\top] \\
\Sigma_{YY} &= \mathrm{E}_{(X,Y)\sim\mathcal{N}(\mu,\Sigma)}[(Y-\mu_Y)(Y-\mu_Y)^\top] \\
\Sigma_{XY} &= \mathrm{E}_{(X,Y)\sim\mathcal{N}(\mu,\Sigma)}[(X-\mu_X)(Y-\mu_Y)^\top] = \Sigma_{YX}^\top \\
\Sigma_{YX} &= \mathrm{E}_{(X,Y)\sim\mathcal{N}(\mu,\Sigma)}[(Y-\mu_Y)(X-\mu_X)^\top] = \Sigma_{XY}^\top
\end{aligned}$$

Partitioned Multivariate Gaussian: Dual Representation

- Precision matrix $$\Gamma = \Sigma^{-1} = \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix}^{-1} = \begin{bmatrix} \Gamma_{XX} & \Gamma_{XY} \\ \Gamma_{YX} & \Gamma_{YY} \end{bmatrix} \quad (1)$$

$$p(\begin{bmatrix} x \\ y \end{bmatrix}; \mu, \Sigma) = \frac{1}{(2\pi)^(n/2)|\Sigma|^{1/2}} \exp\left(-\frac{1}{2} \left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix} \right)^\top \begin{bmatrix} \Gamma_{XX} & \Gamma_{XY} \\ \Gamma_{YX} & \Gamma_{YY} \end{bmatrix} \left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix} \right) \right)$$

- Straightforward to verify from (1) that:

$$\begin{aligned}
\Sigma_{XX} &= \left(\Gamma_{XX} - \Gamma_{XY}\Gamma_{YY}^{-1}\Gamma_{YX}\right)^{-1} \\
\Sigma_{YY} &= \left(\Gamma_{YY} - \Gamma_{YX}\Gamma_{XX}^{-1}\Gamma_{XY}\right)^{-1} \\
\Sigma_{XY} &= -\Gamma_{XX}^{-1}\Gamma_{XY}\left(\Gamma_{YY} - \Gamma_{YX}\Gamma_{XX}^{-1}\Gamma_{XY}\right)^{-1} = \Sigma_{YX}^\top \\
\Sigma_{YX} &= -\Gamma_{YY}^{-1}\Gamma_{YX}\left(\Gamma_{XX} - \Gamma_{XY}\Gamma_{YY}^{-1}\Gamma_{YX}\right)^{-1} = \Sigma_{XY}^\top
\end{aligned}$$

- And swapping the roles of Sigma and Gamma:

$$\begin{aligned}
\Gamma_{XX} &= \left(\Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\right)^{-1} \\
\Gamma_{YY} &= \left(\Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\right)^{-1} \\
\Gamma_{XY} &= -\Sigma_{XX}^{-1}\Sigma_{XY}\left(\Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\right)^{-1} = \Gamma_{YX}^\top \\
\Gamma_{YX} &= -\Sigma_{YY}^{-1}\Sigma_{YX}\left(\Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\right)^{-1} = \Gamma_{XY}^\top
\end{aligned}$$

Marginalization: p(x) = ?

$$p(\begin{bmatrix} x \\ y \end{bmatrix}; \mu, \Sigma) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}\right)^\top \begin{bmatrix} \Gamma_{XX} & \Gamma_{XY} \\ \Gamma_{YX} & \Gamma_{YY} \end{bmatrix} \left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}\right)\right)$$

We integrate out over y to find the marginal:

$$\begin{aligned}
p(x) &= \int p(\begin{bmatrix} x \\ y \end{bmatrix}; \mu, \Sigma) dy \\
&= \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \int \exp\left(-\frac{1}{2}\left((x-\mu_X)^\top \Gamma_{XX}(x-\mu_X) + (y-\mu_Y)^\top \Gamma_{YY}(y-\mu_Y) + 2(y-\mu_Y)^\top \Gamma_{YX}(x-\mu_X)\right)\right) dy \\
&= \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \int \exp\left(-\frac{1}{2}\left((x-\mu_X)^\top \Gamma_{XX}(x-\mu_X) + (y-\mu_Y)^\top \Gamma_{YY}(y-\mu_Y) + 2(y-\mu_Y)^\top \Gamma_{YY}\Gamma_{YY}^{-1}\Gamma_{YX}(x-\mu_X) + (x-\mu_X)^\top \Gamma_{XY}\Gamma_{YY}^{-1}\Gamma_{YY}\Gamma_{YY}^{-1}\Gamma_{YX}(x-\mu_X) - (x-\mu_X)^\top \Gamma_{XY}\Gamma_{YY}^{-1}\Gamma_{YY}\Gamma_{YY}^{-1}\Gamma_{YX}(x-\mu_X)\right)\right) dy \\
&= \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\left((x-\mu_X)^\top \Gamma_{XX}(x-\mu_X) - (x-\mu_X)^\top \Gamma_{XY}\Gamma_{YY}^{-1}\Gamma_{YY}\Gamma_{YY}^{-1}\Gamma_{YX}(x-\mu_X)\right)\right) \int \exp\left(-\frac{1}{2}\left((y-\mu_Y)^\top \Gamma_{YY}(y-\mu_Y) + 2(y-\mu_Y)^\top \Gamma_{YY}\Gamma_{YY}^{-1}\Gamma_{YX}(x-\mu_X) + (x-\mu_X)^\top \Gamma_{XY}\Gamma_{YY}^{-1}\Gamma_{YY}\Gamma_{YY}^{-1}\Gamma_{YX}(x-\mu_X)\right)\right) dy \\
&= \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\left((x-\mu_X)^\top \Gamma_{XX}(x-\mu_X) - (x-\mu_X)^\top \Gamma_{XY}\Gamma_{YY}^{-1}\Gamma_{YX}(x-\mu_X)\right)\right) \int \exp\left(-\frac{1}{2}\left((y-\mu_Y+\Gamma_{YY}^{-1}\Gamma_{YX}(x-\mu_X))^\top \Gamma_{YY}(y-\mu_Y+\Gamma_{YY}^{-1}\Gamma_{YX}(x-\mu_X))\right)\right) dy \\
&= \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\left((x-\mu_X)^\top \Gamma_{XX}(x-\mu_X) - (x-\mu_X)^\top \Gamma_{XY}\Gamma_{YY}^{-1}\Gamma_{YX}(x-\mu_X)\right)\right) (2\pi)^{n_Y/2}|\Gamma_{YY}^{-1}|^{1/2} \\
&= \frac{(2\pi)^{n_Y/2}|\Gamma_{YY}^{-1}|^{1/2}}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\left((x-\mu_X)^\top \Gamma_{XX}(x-\mu_X) - (x-\mu_X)^\top \Gamma_{XY}\Gamma_{YY}^{-1}\Gamma_{YX}(x-\mu_X)\right)\right) \\
&= \frac{(2\pi)^{n_Y/2}|\Gamma_{YY}^{-1}|^{1/2}}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\left((x-\mu_X)^\top \left(\Gamma_{XX} - \Gamma_{XY}\Gamma_{YY}^{-1}\Gamma_{YX}\right)(x-\mu_X)\right)\right)
\end{aligned}$$

Hence we have:

$$X \sim \mathcal{N}(\mu_X, (\Gamma_{XX} - \Gamma_{XY}\Gamma_{YY}^{-1}\Gamma_{YX})^{-1}) = \mathcal{N}(\mu_X, \Sigma_{XX})$$

Note: ***if** we had known beforehand* that p(x) would be a Gaussian distribution, then we could have found the result more quickly. We would have just needed to find $\mu_X = \mathrm{E}[X]$ and $\Sigma_{XX} = \mathrm{E}[(X-\mu_X)(X-\mu_X)^\top]$, which we had available through $\mathcal{N}(\mu, \Sigma)$

Marginalization Recap

If

$$(X, Y) \sim \mathcal{N}(\mu, \Sigma) = \mathcal{N}\left(\begin{bmatrix}\mu_X \\ \mu_Y\end{bmatrix}, \begin{bmatrix}\Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY}\end{bmatrix}\right)$$

Then

$$\begin{aligned} X &\sim \mathcal{N}(\mu_X, \Sigma_{XX}) \\ Y &\sim \mathcal{N}(\mu_Y, \Sigma_{YY}) \end{aligned}$$

Self-quiz

Test your understanding of the completion of squares trick! Let $A \in \mathbf{R}^{n \times n}$ be a positive definite matrix, $b \in \mathbf{R}^n$, and $c \in \mathbf{R}$. Prove that

$$\int_{x \in \mathbf{R}^n} \exp\left(-\frac{1}{2}x^T A x - x^T b - c\right) dx$$

$$= \frac{(2\pi)^{n/2}}{|A|^{1/2} \exp(c - \frac{1}{2} b^T A^{-1} b)}.$$

Conditioning: $p(x \mid Y = y_0) = ?$

$$p(\begin{bmatrix} x \\ y \end{bmatrix}; \mu, \Sigma) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2} \left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix} \right)^\top \begin{bmatrix} \Gamma_{XX} & \Gamma_{XY} \\ \Gamma_{YX} & \Gamma_{YY} \end{bmatrix} \left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix} \right) \right)$$

We have

$$\begin{aligned}
p(x|Y=y_0) &\propto p(\begin{bmatrix} x \\ y_0 \end{bmatrix}; \mu, \Sigma) \\
&\propto \exp\left(-\frac{1}{2}(x-\mu_X)^\top \Gamma_{XX}(x-\mu_X) - (x-\mu_X)^\top \Gamma_{XY}(y_0-\mu_Y) - \frac{1}{2}(y_0-\mu_Y)^\top \Gamma_{YY}(y_0-\mu_Y) \right) \\
&\propto \exp\left(-\frac{1}{2}(x-\mu_X)^\top \Gamma_{XX}(x-\mu_X) - (x-\mu_X)^\top \Gamma_{XY}(y_0-\mu_Y) \right) \\
&= \exp\left(-\frac{1}{2}(x-\mu_X)^\top \Gamma_{XX}(x-\mu_X) - (x-\mu_X)^\top \Gamma_{XX}\Gamma_{XX}^{-1}\Gamma_{XY}(y_0-\mu_Y) - \frac{1}{2}(y_0-\mu_Y)\Gamma_{YX}\Gamma_{XX}^{-1}\Gamma_{XX}\Gamma_{XX}^{-1}\Gamma_{XY}(y_0-\mu_Y) + \frac{1}{2}(y_0-\mu_Y)\Gamma_{YX}\Gamma_{XX}^{-1}\Gamma_{XX}\Gamma_{XX}^{-1}\Gamma_{XY}(y_0-\mu_Y) \right) \\
&= \exp\left(-\frac{1}{2}(x-\mu_X+\Gamma_{XX}^{-1}\Gamma_{XY}(y_0-\mu_Y))^\top \Gamma_{XX}(x-\mu_X+\Gamma_{XX}^{-1}\Gamma_{XY}(y_0-\mu_Y) \right) \exp\left(\frac{1}{2}(y_0-\mu_Y)\Gamma_{YX}\Gamma_{XX}^{-1}\Gamma_{XX}\Gamma_{XX}^{-1}\Gamma_{XY}(y_0-\mu_Y) \right) \\
&\propto \exp\left(-\frac{1}{2}(x-\mu_X+\Gamma_{XX}^{-1}\Gamma_{XY}(y_0-\mu_Y))^\top \Gamma_{XX}(x-\mu_X+\Gamma_{XX}^{-1}\Gamma_{XY}(y_0-\mu_Y) \right)
\end{aligned}$$

Hence we have:

$$\begin{aligned}
X|Y=y_0 &\sim \mathcal{N}(\mu_X - \Gamma_{XX}^{-1}\Gamma_{XY}(y_0-\mu_Y), \Gamma_{XX}^{-1}) \\
&= \mathcal{N}(\mu_X + \Sigma_{XY}\Sigma_{YY}^{-1}(y_0-\mu_Y), \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX})
\end{aligned}$$

- Conditional mean moved according to correlation and variance on measurement
- Conditional covariance does not depend on y_0

Conditioning Recap

If

$$(X, Y) \sim \mathcal{N}(\mu, \Sigma) = \mathcal{N}\left(\begin{bmatrix}\mu_X \\ \mu_Y\end{bmatrix}, \begin{bmatrix}\Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY}\end{bmatrix}\right)$$

Then

$$\begin{aligned} X|Y = y_0 &\sim \mathcal{N}(\mu_X + \Sigma_{XY}\Sigma_{YY}^{-1}(y_0 - \mu_Y), \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}) \\ Y|X = x_0 &\sim \mathcal{N}(\mu_Y + \Sigma_{YX}\Sigma_{XX}^{-1}(x_0 - \mu_X), \Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}) \end{aligned}$$

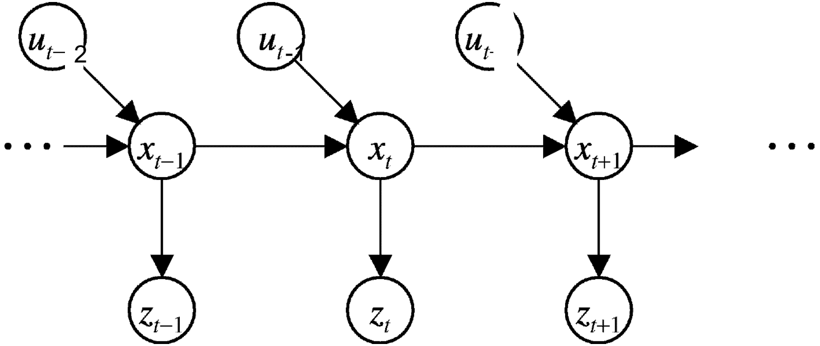
Outline

- Gaussians
- ***Kalman filtering***
- Extend Kalman Filter (EKF)
- Unscented Kalman Filter (UKF) [aka “sigma-point filter”]

Kalman Filter

- Kalman Filter = special case of a Bayes' filter with dynamics and sensory models linear Gaussians:

$$X_0 \sim \mathcal{N}(\mu_0, \Sigma_0)$$

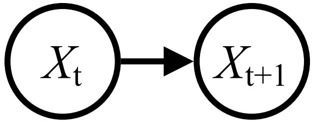


$$\begin{aligned} X_{t+1} &= A_t X_t + B_t u_t + \varepsilon_t \quad \varepsilon_t \sim \mathcal{N}(0, Q_t) \\ Z_t &= C_t X_t + d_t + \delta_t \quad \delta_t \sim \mathcal{N}(0, R_t) \end{aligned}$$

Time update

- Assume we have current belief for $X_{t|0:t}$:

$$p(x_t|z_{0:t}, u_{0:t})$$



- Then, after one time step passes:

$$p(x_{t+1}|z_{0:t}, u_{0:t}) = \int_{x_t} p(x_{t+1}, x_t|z_{0:t}, u_{0:t}) dx_t$$

$$\begin{aligned} p(x_{t+1}, x_t|z_{0:t}, u_{0:t}) &= p(x_{t+1}|x_t, z_{0:t}, u_{0:t}) p(x_t|z_{0:t}, u_{0:t}) \\ &= p(x_{t+1}|x_t, u_t) p(x_t|z_{0:t}, u_{0:t}) \end{aligned}$$

Time Update: Finding the joint $p(x_{t+1}, x_t | z_{0:t}, u_{0:t})$

$$
\begin{aligned}
p(x_{t+1}, x_t | z_{0:t}, u_{0:t}) &= p(x_{t+1} | x_t, u_t) p(x_t | z_{0:t}, u_{0:t}) \\
&= \frac{1}{(2\pi)^{n/2} |\Sigma_{t|0:t}|^{1/2}} e^{-\frac{1}{2}(x_t - \mu_{t|0:t})^\top \Sigma_{t|0:t}^{-1} (x_t - \mu_{t|0:t})} \\
&\quad \frac{1}{(2\pi)^{n/2} |Q_t|^{1/2}} e^{-\frac{1}{2}(x_{t+1} - (A_t x_t + B_t u_t))^\top Q_t^{-1} (x_{t+1} - (A_t x_t + B_t u_t))}
\end{aligned}
$$

- Now we can choose to continue by either of
  - (i) mold it into a standard multivariate Gaussian format so we can read of the joint distribution’s mean and covariance
  - (ii) observe this is a quadratic form in x_{t} and x_{t+1} in the exponent; the exponent is the only place they appear; hence we know this is a multivariate Gaussian. We directly compute its mean and covariance. [usually simpler!]

Time Update: Finding the joint $p(x_{t+1}, x_t | z_{0:t}, u_{0:t})$

- We follow (ii) and find the means and covariance matrices in

$$(X_{t+1}, X_t)|z_{0:t}, u_{0:t} \sim \mathcal{N}\left(\begin{bmatrix}\mu_{t|0:t} \\ \mu_{t+1|0:t}\end{bmatrix}, \begin{bmatrix}\Sigma_{t|0:t} & \Sigma_{t,t+1|0:t} \\ \Sigma_{t+1,t|0:t} & \Sigma_{t+1|0:t}\end{bmatrix}\right)$$

$\mu_{t|0:t}$ and $\Sigma_{t|0:t}$ are available from previous time step

$$\begin{aligned}
\mu_{t+1|0:t} &= \mathrm{E}[X_{t+1}|z_{0:t}, u_{0:t}] \\
&= \mathrm{E}[A_t X_t + B_t u_t + \epsilon_t | z_{0:t}, u_{0:t}] \\
&= A_t \mathrm{E}[X_t | z_{0:t}, u_{0:t}] + B_t u_t + \mathrm{E}[\epsilon_t | z_{0:t}, u_{0:t}] \\
&= A_t \mu_{t|0:t} + B_t u_t
\end{aligned}$$

$$\begin{aligned}
\mu_{t+1|0:t} &= \mathrm{E}[X_{t+1|0:t}] \\
&= \mathrm{E}[A_t X_{t|0:t} + B_t u_t + \epsilon_{t|0:t}] \\
&= A_t \mathrm{E}[X_{t|0:t}] + B_t u_t + \mathrm{E}[\epsilon_{t|0:t}] \\
&= A_t \mu_{t:0:t} + B_t u_t
\end{aligned}$$

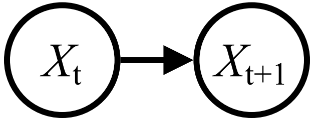
$$\begin{aligned}
\Sigma_{t+1|0:t} &= \mathrm{E}[(X_{t+1|0:t} - \mu_{t+1|0:t})(X_{t+1|0:t} - \mu_{t+1|0:t})^\top] \\
&= \mathrm{E}[((A_t X_{t|0:t} + B_t u_t + \epsilon_t) - (A_t \mu_{t|0:t} + B_t u_t))((A_t X_{t|0:t} + B_t u_t + \epsilon_t) - (A_t \mu_{t|0:t} + B_t u_t))^\top] \\
&= \mathrm{E}[(A_t (X_{t|0:t} - \mu_{t|0:t}) + \epsilon_t)(A_t (X_{t|0:t} - \mu_{t|0:t}) + \epsilon_t)^\top] \\
&= \mathrm{E}[A_t (X_{t|0:t} - \mu_{t|0:t})(X_{t|0:t} - \mu_{t|0:t})^\top A^\top] + \mathrm{E}[\epsilon_t (A_t (X_{t|0:t} - \mu_{t|0:t}))^\top] + \mathrm{E}[(A_t (X_{t|0:t} - \mu_{t|0:t}))\epsilon_t^\top] + \mathrm{E}[\epsilon_t \epsilon_t^\top] \\
&= A_t \mathrm{E}[(X_{t|0:t} - \mu_{t|0:t})(X_{t|0:t} - \mu_{t|0:t})^\top] A_t^\top + \mathrm{E}[\epsilon_t]\mathrm{E}[(A_t (X_{t|0:t} - \mu_{t|0:t}))^\top] + \mathrm{E}[A_t (X_{t|0:t} - \mu_{t|0:t})]\mathrm{E}[\epsilon_t] + \mathrm{E}[\epsilon_t \epsilon_t^\top] \\
&= A_t \Sigma_{t|0:t} A_t^\top + 0 + 0 + Q_t
\end{aligned}$$

$$\begin{aligned}
\Sigma_{t,t+1|0:t} &= \mathrm{E}[(X_{t|0:t} - \mu_{t|0:t})(X_{t+1|0:t} - \mu_{t+1|0:t})^\top] \\
&= \Sigma_{t|0:t} A_t^\top
\end{aligned}$$

[Exercise: Try to prove each of these without referring to this slide!]

Time Update Recap

- Assume we have



$$\begin{aligned} X_{t|0:t} &\sim \mathcal{N}(\mu_{t|0:t}, \Sigma_{t|0:t}) \\ X_{t+1} &= A_t X_t + B_t u_t + \epsilon_t, \\ \epsilon_t &\sim \mathcal{N}(0, Q_t), \text{ and independent of } x_{0:t}, z_{0:t}, u_{0:t}, \epsilon_{0:t-1} \end{aligned}$$

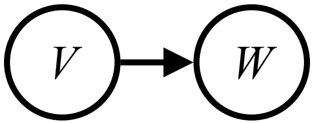
- Then we have

$$\begin{aligned} (X_{t|0:t}, X_{t+1|0:t}) &\sim \mathcal{N}\left(\begin{bmatrix} \mu_{t|0:t} \\ \mu_{t+1|0:t} \end{bmatrix}, \begin{bmatrix} \Sigma_{t|0:t} & \Sigma_{t,t+1|0:t} \\ \Sigma_{t+1,t|0:t} & \Sigma_{t+1|0:t} \end{bmatrix}\right) \\ &= \mathcal{N}\left(\begin{bmatrix} \mu_{t|0:t} \\ A_t \mu_{t|0:t} + B_t u_t \end{bmatrix}, \begin{bmatrix} \Sigma_{t|0:t} & \Sigma_{t|0:t} A_t^\top \\ A_t \Sigma_{t|0:t} & A_t \Sigma_{t|0:t} A_t^\top + Q_t \end{bmatrix}\right) \end{aligned}$$

- Marginalizing the joint, we immediately get

$$X_{t+1|0:t} \sim \mathcal{N}\left(A_t \mu_{t|0:t} + B_t u_t, A_t \Sigma_{t|0:t} A_t^\top + Q_t\right)$$

Generality!



- Assume we have

$$
\begin{aligned}
V &\sim \mathcal{N}(\mu_V, \Sigma_{VV}) \\
W &= AV + b + \epsilon, \\
\epsilon &\sim \mathcal{N}(0, Q), \text{ and independent of } V
\end{aligned}
$$

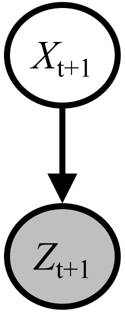
- Then we have

$$
\begin{aligned}
(V, W) &\sim \mathcal{N}\left(\begin{bmatrix} \mu_V \\ \mu_W \end{bmatrix}, \begin{bmatrix} \Sigma_{VV} & \Sigma_{VW} \\ \Sigma_{WV} & \Sigma_{WW} \end{bmatrix}\right) \\
&= \mathcal{N}\left(\begin{bmatrix} \mu_V \\ A_t\mu_V + b \end{bmatrix}, \begin{bmatrix} \Sigma_{VV} & \Sigma_{VV}A^\top \\ A\Sigma_{VV} & A\Sigma_{VV}A^\top + Q \end{bmatrix}\right)
\end{aligned}
$$

- Marginalizing the joint, we immediately get

$$
W \sim \mathcal{N}\left(A\mu_V + v, A\Sigma_{VV}A^\top + Q\right)
$$

Observation update



- Assume we have:

$$\begin{aligned} X_{t+1|0:t} &\sim \mathcal{N}\left(\mu_{t+1|0:t}, \Sigma_{t+1|0:t}\right) \\ Z_{t+1} &\sim C_{t+1} X_{t+1} + d_{t+1} + \delta_{t+1} \\ \delta_{t+1} &\sim \mathcal{N}(0, R_t), \text{ and independent of } x_{0:t+1}, z_{0:t}, u_{0:t}, \epsilon_{0:t}, \end{aligned}$$

- Then:

$$(X_{t+1|0:t}, Z_{t+1|0:t}) \sim \mathcal{N}\left(\begin{bmatrix} \mu_{t+1|0:t} \\ C_{t+1}\mu_{t+1|0:t} + d \end{bmatrix}, \begin{bmatrix} \Sigma_{t+1|0:t} & \Sigma_{t+1|0:t} C_{t+1}^\top \\ C_{t+1}\Sigma_{t+1|0:t} & C_{t+1}\Sigma_{t+1|0:t} C_{t+1}^\top + R_{t+1} \end{bmatrix}\right)$$

- And, by conditioning on $Z_{t+1} = z_{t+1}$ (see lecture slides on Gaussians) we readily get:

$$\begin{aligned} X_{t+1} | z_{0:t+1}, u_{0:t} &= X_{t+1|0:t+1} \\ &\sim \mathcal{N}\left(\mu_{t+1|0:t} + \Sigma_{t+1|0:t} C_{t+1}^\top (C_{t+1}\Sigma_{t+1|0:t} C_{t+1}^\top + R_{t+1})^{-1}(z_{t+1} - (C_{t+1}\mu_{t+1|0:t} + d)),\right. \\ &\qquad \left.\Sigma_{t+1|0:t} - \Sigma_{t+1|0:t} C_{t+1}^\top (C_{t+1}\Sigma_{t+1|0:t} C_{t+1}^\top + R_{t+1})^{-1} C_{t+1}\Sigma_{t+1|0:t}\right) \end{aligned}$$

Complete Kalman Filtering Algorithm

- At time 0: $X_0 \sim \mathcal{N}(\mu_{0|0}, \Sigma_{0|0})$
- For t = 1, 2, …
  - Dynamics update:

$$\begin{aligned}\mu_{t+1|0:t} &= A_t \mu_{t|0:t} + B_t u_t \\ \Sigma_{t+1|0:t} &= A_t \Sigma_{t|0:t} A_t^\top + Q_t\end{aligned}$$

  - Measurement update:

$$\begin{aligned}\mu_{t+1|0:t+1} &= \mu_{t+1|0:t} + \Sigma_{t+1|0:t} C_{t+1}^\top (C_{t+1} \Sigma_{t+1|0:t} C_{t+1}^\top + R_{t+1})^{-1} (z_{t+1} - (C_{t+1} \mu_{t+1|0:t} + d)) \\ \Sigma_{t+1|0:t+1} &= \Sigma_{t+1|0:t} - \Sigma_{t+1|0:t} C_{t+1}^\top (C_{t+1} \Sigma_{t+1|0:t} C_{t+1}^\top + R_{t+1})^{-1} C_{t+1} \Sigma_{t+1|0:t}\end{aligned}$$

    - Often written as:

$$\begin{aligned}K_{t+1} &= \Sigma_{t+1|0:t} C_{t+1}^\top (C_{t+1} \Sigma_{t+1|0:t} C_{t+1}^\top + R_{t+1})^{-1} \\ \mu_{t+1|0:t+1} &= \mu_{t+1|0:t} + K_{t+1} (z_{t+1} - (C_{t+1} \mu_{t+1|0:t} + d)) \\ \Sigma_{t+1|0:t+1} &= (I - K_{t+1} C_{t+1}) \Sigma_{t+1|0:t}\end{aligned}$$

(Kalman gain)

"innovation"

Kalman Filter Summary

- Highly efficient: Polynomial in measurement dimensionality k and state dimensionality n:

$$O(k^{2.376} + n^2)$$

- Optimal for linear Gaussian systems!

Outline

- Gaussians
- Kalman filtering
- ***Extend Kalman Filter (EKF)***
- Unscented Kalman Filter (UKF) [aka “sigma-point filter”]

Nonlinear Dynamical Systems

- Most realistic robotic problems involve nonlinear functions:

$$X_{t+1} = f_t(X_t, u_t) + \varepsilon_t \quad \varepsilon_t \sim \mathcal{N}(0, Q_t)$$

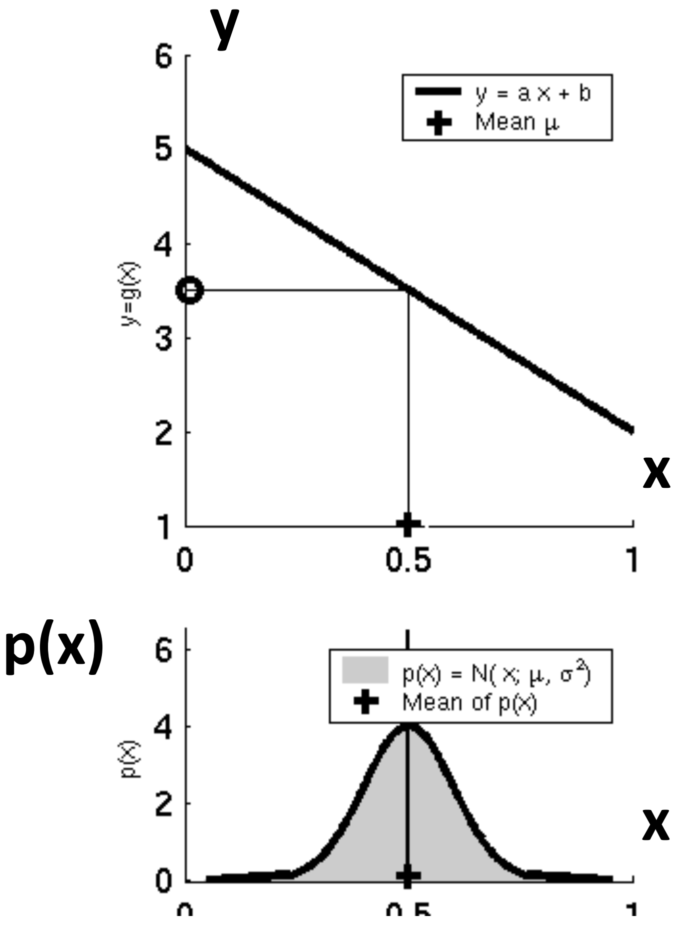
$$Z_t = h_t(X_t) + \delta_t \quad \delta_t \sim \mathcal{N}(0, R_t)$$

  - Versus linear setting:

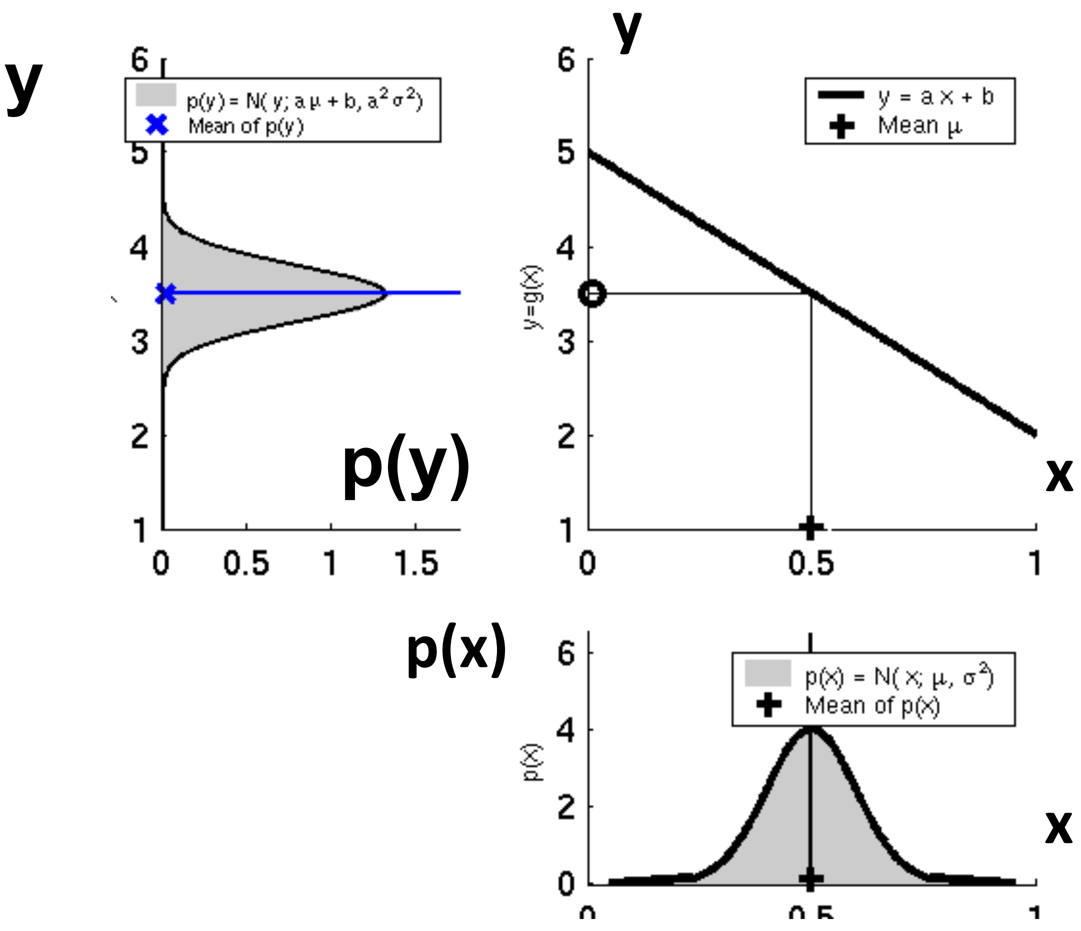
$$X_{t+1} = A_t X_t + B_t u_t + \varepsilon_t \quad \varepsilon_t \sim \mathcal{N}(0, Q_t)$$

$$Z_t = C_t X_t + d_t + \delta_t \quad \delta_t \sim \mathcal{N}(0, R_t)$$

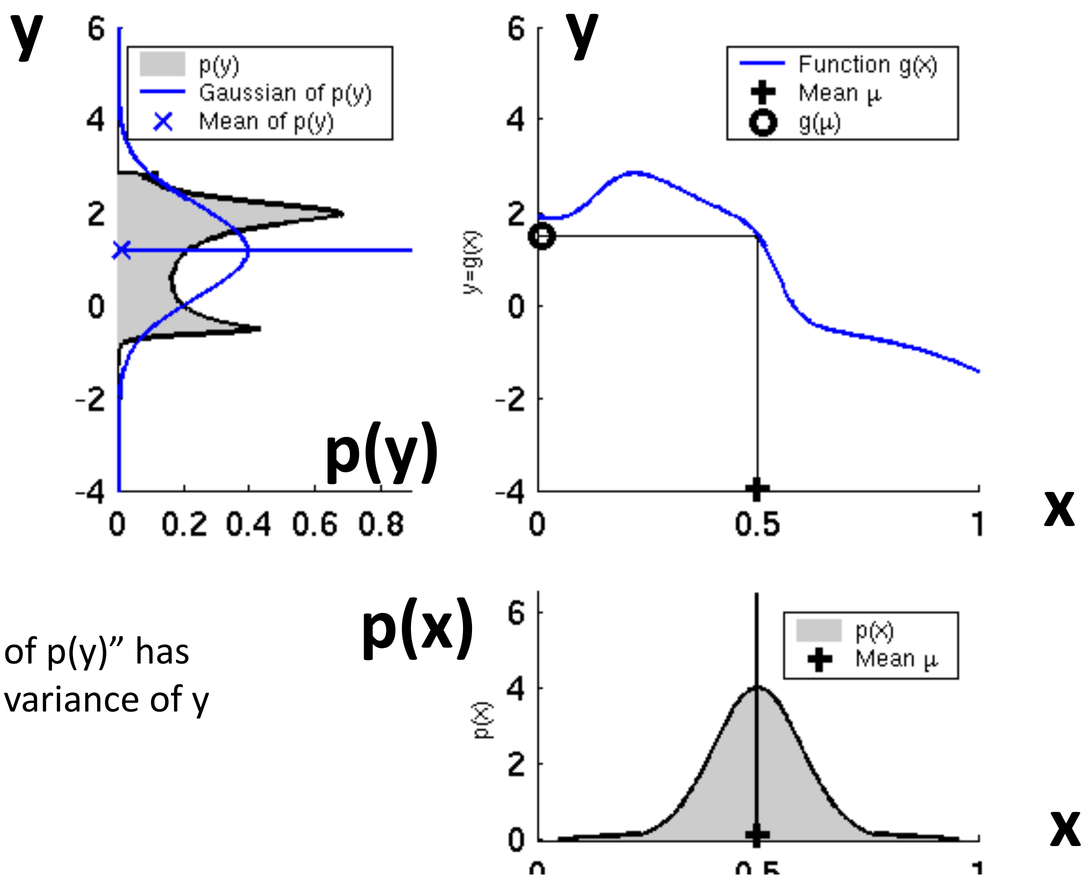
Linearity Assumption Revisited



Linearity Assumption Revisited

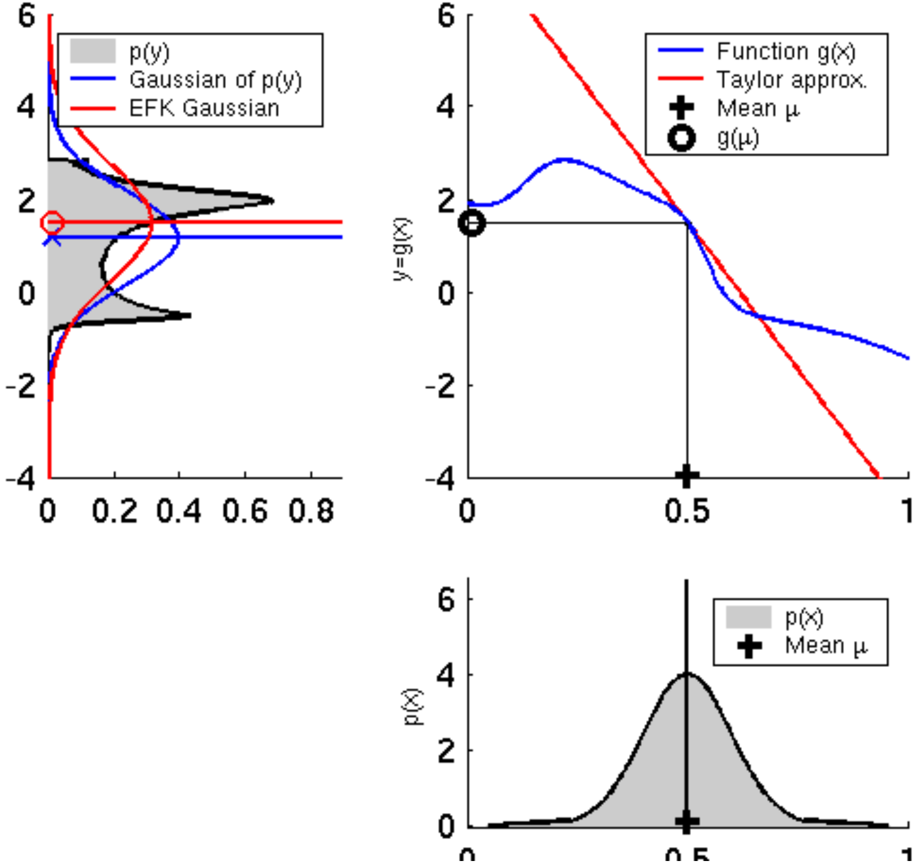


Non-linear Function

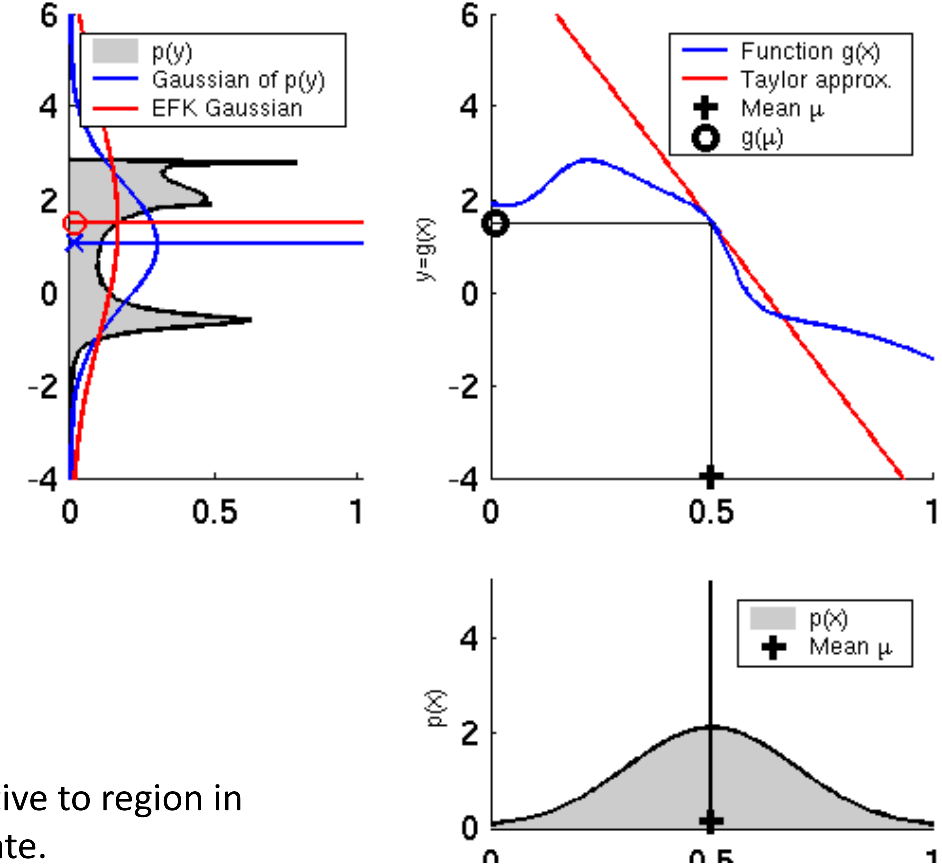


"Gaussian of p(y)" has mean and variance of y under p(y)

EKF Linearization (1)

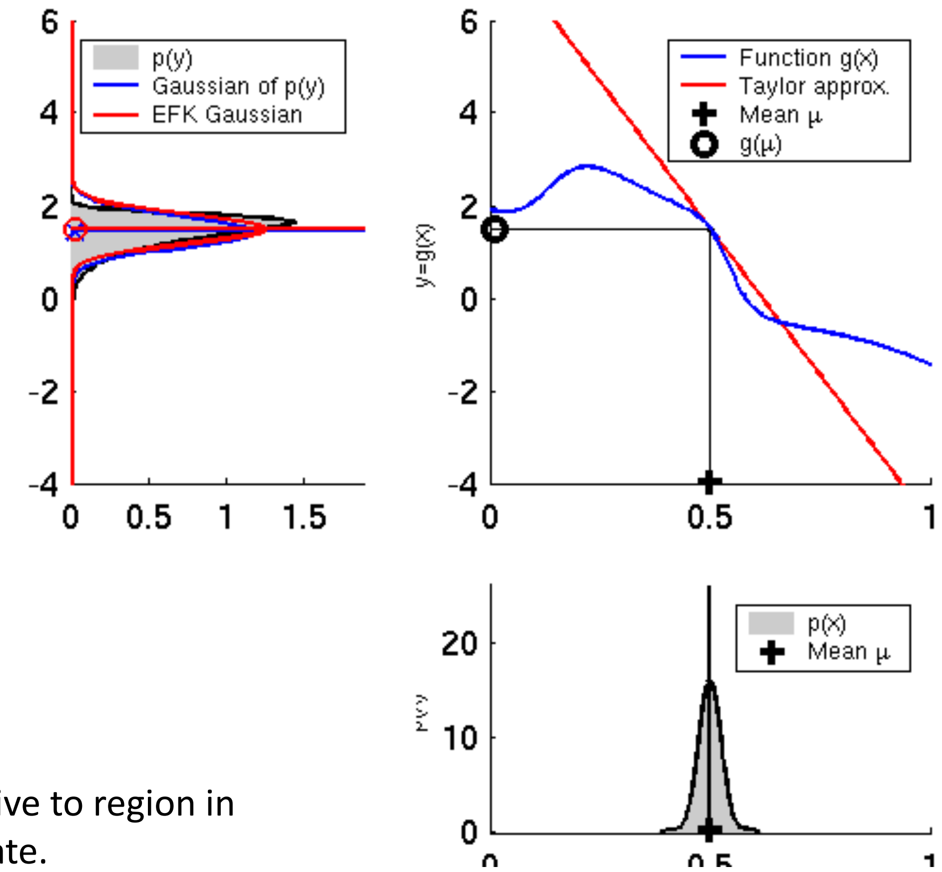


EKF Linearization (2)



p(x) has ***HIGH*** variance relative to region in which linearization is accurate.

EKF Linearization (3)



p(x) has ***LOW*** variance relative to region in which linearization is accurate.

EKF Linearization: First Order Taylor Series Expansion

- **Dynamics model:** for x_t "close to" μ_t we have:

$$
\begin{aligned}
f_t(x_t, u_t) &\approx f_t(\mu_t, u_t) + \frac{\partial f_t(\mu_t, u_t)}{\partial x_t}(x_t - \mu_t) \\
&= f_t(\mu_t, u_t) + F_t(x_t - \mu_t)
\end{aligned}
$$

- **Measurement model:** for x_t "close to" μ_t we have:

$$
\begin{aligned}
h_t(x_t) &\approx h_t(\mu_t) + \frac{\partial h_t(\mu_t)}{\partial x_t}(x_t - \mu_t) \\
&= h_t(\mu_t) + H_t(x_t - \mu_t)
\end{aligned}
$$

EKF Algorithm

- At time 0: $X_0 \sim \mathcal{N}(\mu_{0|0}, \Sigma_{0|0})$
- For t = 1, 2, …
  - Dynamics update: $f_t(x_t, u_t) \quad \approx \quad a_{0,t} + F_t(x_t - \mu_{t|0:t})$

$$
\begin{aligned}
(a_{0,t}, F_t) &= \text{linearize}(f_t, \mu_{t|0:t}, \Sigma_{t|0:t}, u_t) \\
\mu_{t+1|0:t} &= a_{0,t} \\
\Sigma_{t+1|0:t} &= F_t \Sigma_{t|0:t} F_t^\top + Q_t
\end{aligned}
$$

  - Measurement update: $h_{t+1}(x_{t+1}) \quad \approx \quad c_{0,t+1} + H_{t+1}(x_{t+1} - \mu_{t+1|0:t})$

$$
\begin{aligned}
(c_{0,t+1}, H_{t+1}) &= \text{linearize}(h_{t+1}, \mu_{t+1|0:t}, \Sigma_{t+1|0:t}) \\
K_{t+1} &= \Sigma_{t+1|0:t} H_{t+1}^\top (H_{t+1} \Sigma_{t+1|0:t} H_{t+1}^\top + R_{t+1})^{-1} \\
\mu_{t+1|0:t+1} &= \mu_{t+1|0:t} + K_{t+1}(z_{t+1} - c_{0,t+1}) \\
\Sigma_{t+1|0:t+1} &= (I - K_{t+1} H_{t+1}) \Sigma_{t+1|0:t}
\end{aligned}
$$

EKF Summary

- Highly efficient: Polynomial in measurement dimensionality k and state dimensionality n:

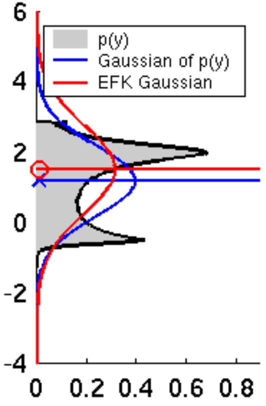
  $$O(k^{2.376} + n^2)$$

- Not optimal!
- Can diverge if nonlinearities are large!
- Works surprisingly well even when all assumptions are violated!

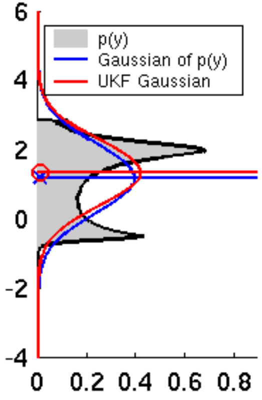
Outline

- Gaussians
- Kalman filtering
- Extend Kalman Filter (EKF)
- ***Unscented Kalman Filter (UKF) [aka "sigma-point filter"]***

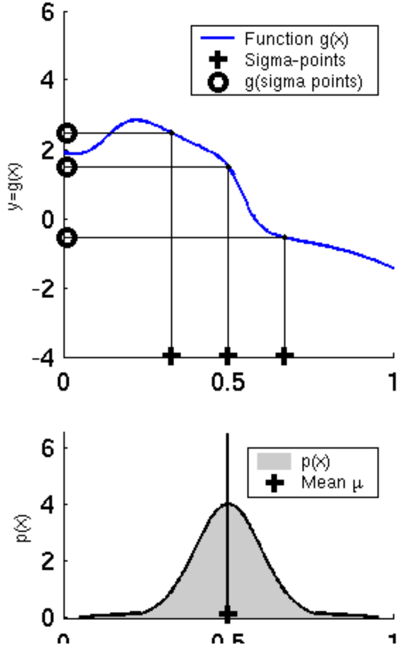
Linearization via Unscented Transform



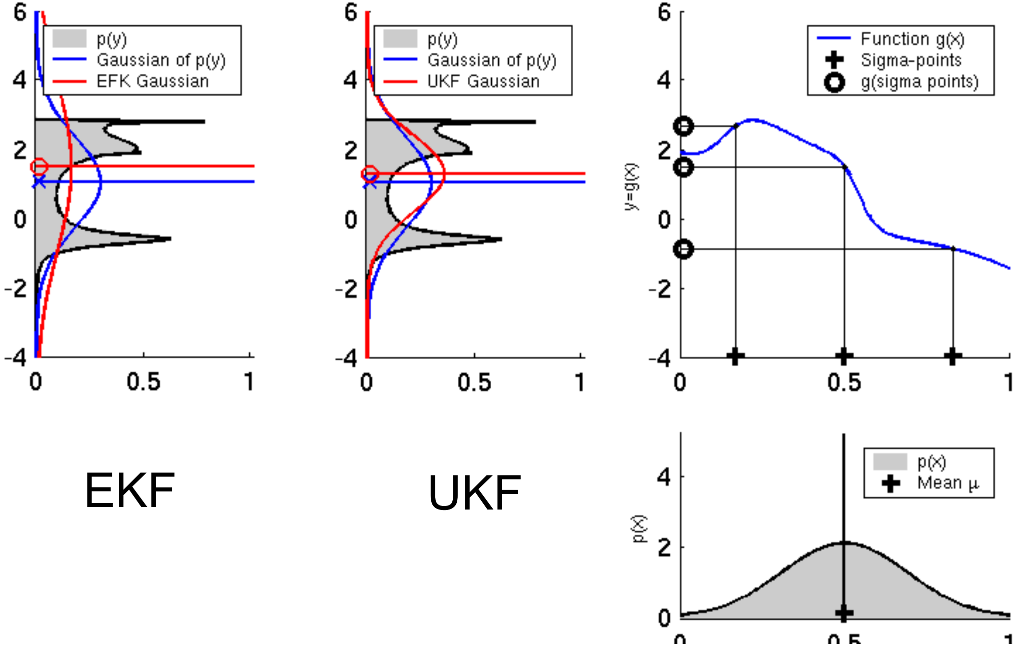
EKF



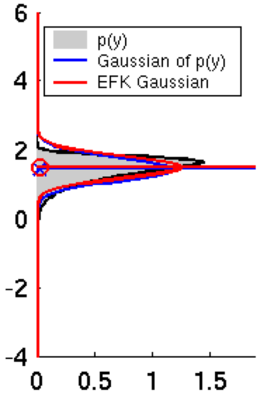
UKF



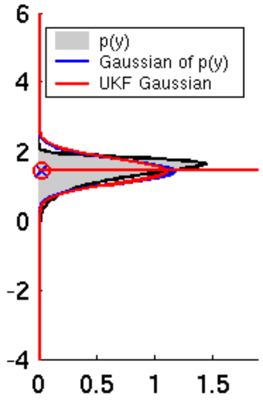
UKF Sigma-Point Estimate (2)



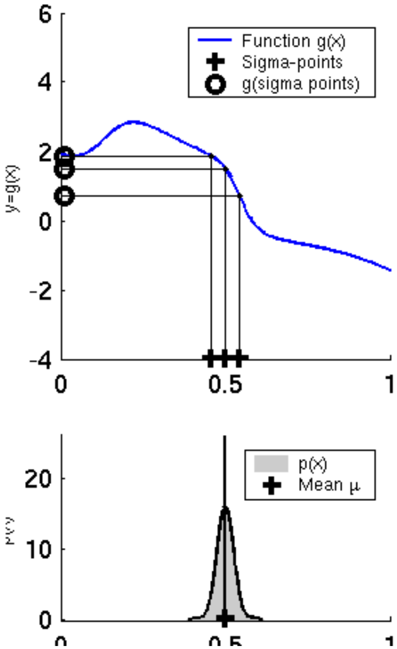
UKF Sigma-Point Estimate (3)



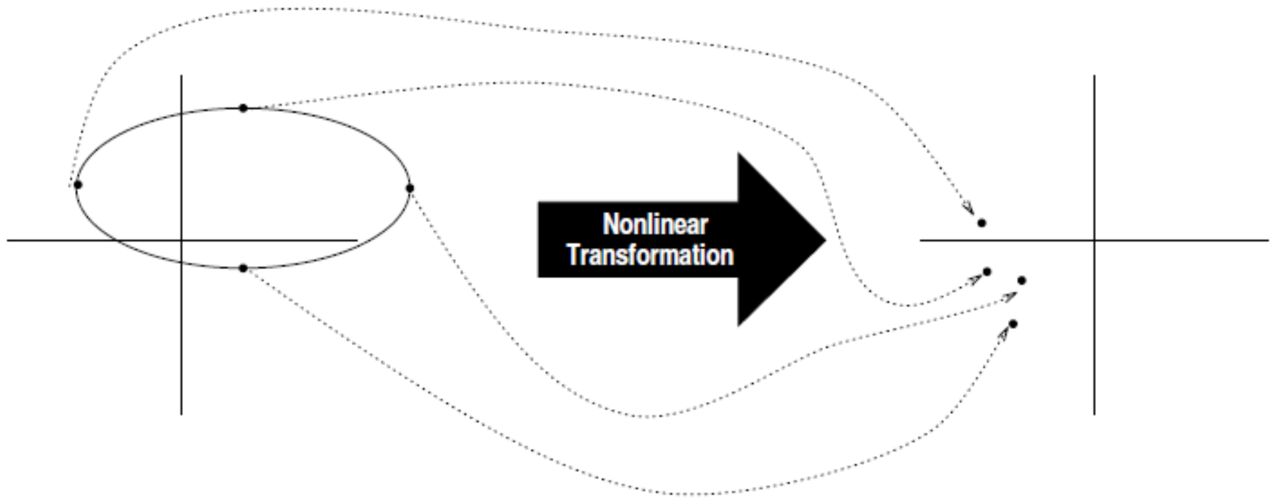
EKF



UKF



UKF Sigma-Point Estimate (4)



[Julier and Uhlmann, 1997]

UKF intuition why it can perform better

- Assume we know the distribution over X and it has a mean \bar{x}
- Y = f(X)

$$\mathbf{f}[\mathbf{x}] = \mathbf{f}[\bar{\mathbf{x}} + \boldsymbol{\delta}\mathbf{x}]$$
$$= \mathbf{f}[\bar{\mathbf{x}}] + \nabla\mathbf{f}\boldsymbol{\delta}\mathbf{x} + \frac{1}{2}\nabla^2\mathbf{f}\boldsymbol{\delta}\mathbf{x}^2 + \frac{1}{3!}\nabla^3\mathbf{f}\boldsymbol{\delta}\mathbf{x}^3 + \frac{1}{4!}\nabla^4\mathbf{f}\boldsymbol{\delta}\mathbf{x}^4 + \cdots$$

$$\bar{\mathbf{y}} = \mathbf{f}[\bar{\mathbf{x}}] + \frac{1}{2}\nabla^2\mathbf{f}\,\mathbf{P}_{xx} + \frac{1}{2}\nabla^4\mathbf{f}\,\mathrm{E}\left[\boldsymbol{\delta}\mathbf{x}^4\right] + \cdots$$

$$\mathbf{P}_{yy} = \nabla\mathbf{f}\,\mathbf{P}_{xx}(\nabla\mathbf{f})^T + \frac{1}{2\times 4!}\nabla^2\mathbf{f}\left(\mathrm{E}\left[\boldsymbol{\delta}\mathbf{x}^4\right] - \mathrm{E}\left[\boldsymbol{\delta}\mathbf{x}^2\mathbf{P}_{yy}\right] - \mathrm{E}\left[\mathbf{P}_{yy}\boldsymbol{\delta}\mathbf{x}^2\right] + \mathbf{P}_{yy}^2\right)(\nabla^2\mathbf{f})^T + \frac{1}{3!}\nabla^3\mathbf{f}\mathrm{E}\left[\boldsymbol{\delta}\mathbf{x}^4\right](\nabla\mathbf{f})^T + \cdots.$$

- EKF approximates f to first order and ignores higher-order terms
- UKF uses f exactly, but approximates p(x).

Original Unscented Transform

- Picks a minimal set of sample points that match 1st, 2nd and 3rd moments of a Gaussian:

$$
\begin{aligned}
\mathcal{X}_0 &= \bar{\mathbf{x}} & W_0 &= \kappa/(n+\kappa) \\
\mathcal{X}_i &= \bar{\mathbf{x}} + \left(\sqrt{(n+\kappa)\mathbf{P}_{xx}}\right)_i & W_i &= 1/2(n+\kappa) \\
\mathcal{X}_{i+n} &= \bar{\mathbf{x}} - \left(\sqrt{(n+\kappa)\mathbf{P}_{xx}}\right)_i & W_{i+n} &= 1/2(n+\kappa)
\end{aligned}
$$

- \bar{x} = mean, P_{xx} = covariance, i → i'th column, x in R^n
- κ : extra degree of freedom to fine-tune the higher order moments of the approximation; when x is Gaussian, n+κ = 3 is a suggested heuristic
- L = \sqrt{P_{xx}} can be chosen to be any matrix satisfying:
  - $L L^T = P_{xx}$

[Julier and Uhlmann, 1997]

Unscented Kalman filter

- Dynamics update:
  - Can simply use unscented transform and estimate the mean and variance at the next time from the sample points
- Observation update:
  - Use sigma-points from unscented transform to compute the covariance matrix between x_t and z_t. Then can do the standard update.

Algorithm Unscented_Kalman_filter($\mu_{t-1}, \Sigma_{t-1}, u_t, z_t$):

1. $\mathcal{X}_{t-1} = \left(\mu_{t-1} \quad \mu_{t-1} + \gamma\sqrt{\Sigma_{t-1}} \quad \mu_{t-1} - \gamma\sqrt{\Sigma_{t-1}}\right)$
2. $\bar{\mathcal{X}}_t^* = g(\mu_t, \mathcal{X}_{t-1})$
3. $\bar{\mu}_t = \sum_{i=0}^{2n} w_m^{[i]} \bar{\mathcal{X}}_t^{*[i]}$
4. $\bar{\Sigma}_t = \sum_{i=0}^{2n} w_c^{[i]} (\bar{\mathcal{X}}_t^{*[i]} - \bar{\mu}_t)(\bar{\mathcal{X}}_t^{*[i]} - \bar{\mu}_t)^\top + R_t$
5. $\bar{\mathcal{X}}_t = \left(\bar{\mu}_t \quad \bar{\mu}_t + \gamma\sqrt{\bar{\Sigma}_t} \quad \bar{\mu}_t - \gamma\sqrt{\bar{\Sigma}_t}\right)$
6. $\bar{\mathcal{Z}}_t = h(\bar{\mathcal{X}}_t)$
7. $\hat{z}_t = \sum_{i=0}^{2n} w_m^{[i]} \bar{\mathcal{Z}}_t^{[i]}$
8. $S_t = \sum_{i=0}^{2n} w_c^{[i]} \left(\bar{\mathcal{Z}}_t^{[i]} - \hat{z}_t\right)\left(\bar{\mathcal{Z}}_t^{[i]} - \hat{z}_t\right)^\top + Q_t$
9. $\bar{\Sigma}_t^{x,z} = \sum_{i=0}^{2n} w_c^{[i]} \left(\bar{\mathcal{X}}_t^{[i]} - \bar{\mu}_t\right)\left(\bar{\mathcal{Z}}_t^{[i]} - \hat{z}_t\right)^\top$
10. $K_t = \bar{\Sigma}_t^{x,z} S_t^{-1}$
11. $\mu_t = \bar{\mu}_t + K_t(z_t - \hat{z}_t)$
12. $\Sigma_t = \bar{\Sigma}_t - K_t S_t K_t^\top$
13. **return** μ_t, Σ_t

Here $L = \sqrt{\Sigma}$ can be chosen to be any $n \times n$ matrix satisfying: $LL^\top = \Sigma$
Technically this is an abuse of notation for the symbol $\sqrt{\ }$.

[Table 3.4 in Probabilistic Robotics]

UKF Summary

- Highly efficient: Same complexity as EKF, with a constant factor slower in typical practical applications
- Better linearization than EKF: Accurate in first two terms of Taylor expansion (EKF only first term) + capturing more aspects of the higher order terms
- Derivative-free: No Jacobians needed
- Still not optimal!

Forthcoming

- How to estimate A_t, B_t, C_t, Q_t, R_t from data ($z_{0:T}$, $u_{0:T}$)
  - EM algorithm
- How to compute $p(x_t | z_{0:T}, u_{0:T})$ (= smoothing) (note the capital "T")

Things to be aware of (but we won't cover)

- Square-root Kalman filter --- keeps track of square root of covariance matrices --- equally fast, numerically more stable (bit more complicated conceptually)
- Very large systems with sparsity structure
  - Sparse Information Filter
- Very large systems with low-rank structure
  - Ensemble Kalman Filter
- Kalman filtering over SE(3)
- How to estimate A_t, B_t, C_t, Q_t, R_t from data ($z_{0:T}$, $u_{0:T}$)
  - EM algorithm
- How to compute $p(x_t | z_{0:T}, u_{0:T})$ (= smoothing) (note the capital "T")

Things to be aware of (but we won't cover)

- If $A_t = A$, $Q_t = Q$, $C_t = C$, $R_t = R$
  - If system is "observable" then covariances and Kalman gain will converge to steady-state values for t -> 1
    - Can take advantage of this: pre-compute them, only track the mean, which is done by multiplying Kalman gain with "innovation"
  - System is observable if and only if the following holds true: if there were zero noise you could determine the initial state after a finite number of time steps
  - Observable if and only if: $\text{rank}([\, C \,;\, CA \,;\, CA^2 \,;\, CA^3 \,;\, \ldots \,;\, CA^{n-1}]) = n$
  - Typically if a system is not observable you will want to add a sensor to make it observable
- Kalman filter can also be derived as the (recursively computed) least-squares solutions to a (growing) set of linear equations

Kalman filter property

- If system is observable (=dual of controllable!) then Kalman filter will converge to the true state.

- System is observable if and only if:

  $O = [C \; ; CA \; ; CA^2 \; ; \ldots \; ; CA^{n-1}]$ is full column rank (1)

Intuition: if no noise, we observe y_0, y_1, … and we have that the unknown initial state x_0 satisfies:

$y_0 = C x_0$

$y_1 = CA x_0$

...

$y_K = CA^K x_0$

This system of equations has a unique solution x_0 iff the matrix $[C; CA; \ldots CA^K]$ has full column rank. B/c any power of a matrix higher than n can be written in terms of lower powers of the same matrix, condition (1) is sufficient to check (i.e., the column rank will not grow anymore after having reached K=n-1).